



**UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS**

**FACULTAD DE CIENCIAS MATEMÁTICAS**

**EAP. DE ESTADÍSTICA**

**Modelo de respuesta aleatorizada : una aplicación a  
estudiantes de la Facultad de Ciencias Matemáticas de la  
Universidad Nacional Mayor de San Marcos**

**TESINA**

Para optar el Título de Licenciada en Estadística

**AUTOR**

**Patricia Gianina Canseco Valdez**

LIMA – PERÚ  
2009

**MODELO DE RESPUESTA ALEATORIZADA: UNA APLICACIÓN A  
ESTUDIANTES DE LA FACULTAD DE CIENCIAS MATEMÁTICAS DE LA  
UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS**

PATRICIA GIANINA CANSECO VALDEZ

Tesina presentada a consideración del Cuerpo Docente de la Facultad de Ciencias Matemáticas, de la Universidad Nacional Mayor de San Marcos, como parte de los requisitos para obtener el Título Profesional de Licenciado en Estadística.

Aprobada por:

Lic. Víctor Ramón García Herbozo .....  
Presidente

Mg. Olga Lidia Solano Dávila .....  
Miembro – Asesor

Lima – Perú  
Diciembre 2009

## **FICHA CATALOGRÁFICA**

CANSECO VALDEZ, PATRICIA GIANINA

Modelo de Respuesta Aleatorizada: una aplicación a estudiantes de la Facultad de Ciencias Matemáticas de la Universidad Nacional Mayor de San Marcos, (Lima) 2009.

ix, 55 p., 29.7 cm. (UNMSM, Licenciada en Estadística, 2009)

Tesina, Universidad Nacional Mayor de San Marcos, Facultad de Ciencias Matemáticas. Estadística.

i. UNMSM / F. de C.M. ii. Título (serie).

*Dedico este trabajo a Dios por darme el mejor ejemplo  
a seguir que es mi madre, a mi hijo que es la razón de  
mi existencia, a toda mi familia en general y a mis  
abuelos que hoy son la luz que ilumina nuestro hogar.*

*Le agradezco a la profesora Olga Solano por sus enseñanzas, no sólo en el ámbito estadístico sino en el espiritual, a mis padres, mi hijo, mi pareja, hermanas, sobrinos, profesores, amigos, compañeros de trabajo y a todas las personas que me ayudan a cada día ser mejor, gracias por los bonitos momentos y las enseñanzas de vida que me dan cada día.*

# **RESUMEN**

## **MODELO DE RESPUESTA ALEATORIZADA: UNA APLICACIÓN A ESTUDIANTES DE LA FACULTAD DE CIENCIAS MATEMÁTICAS DE LA UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS**

**PATRICIA GIANINA CANSECO VALDEZ**

DICIEMBRE – 2009

ASESORA : Mg. Olga Lidia Solano Dávila

TÍTULO OBTENIDO : Licenciado en Estadística

Este trabajo de investigación presenta la teoría y la aplicación del Modelo de Respuesta Aleatorizada.

EL Modelo de Respuesta Aleatorizada Estratificada propuesto por Kim y Elam (2003) se aplicó en una encuesta realizada en la Facultad de Ciencias Matemáticas, para investigar el comportamiento de los jóvenes con respecto a las drogas, sexo, alcohol y hurto, estos resultados fueron comparados con el Método Convencional de entrevista directa. La población en estudio comprende a los Alumnos Matriculados en el Semestre 2009-II.

**PALABRAS CLAVES:** MODELOS DE RESPUESTA ALEATORIZADA  
MECANISMO ALEATORIZADO  
MODELOS DE RESPUESTA ALEATORIZADA  
ESTRATIFICADA  
PREGUNTA DELICADA

## **SUMMARY**

### **RANDOMIZED RESPONSE MODEL: APPLICATION TO STUDENTS OF THE FACULTY OF MATHEMATICAL SCIENCES OF THE UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS**

**PATRICIA GIANINA CANSECO VALDEZ**

DECEMBER – 2009

ADVISORY : M.Sc. Olga Lidia Solano Dávila

OBTAINED TITLE : Licenciante in Statistic

This work of research presents the theory and application of the Randomized Response Model.

The Stratified Randomized Response Model proposed by Kim and Elam (2003) was applied in a survey conducted in the Faculty of Mathematical Sciences, to investigate the behavior of young people towards drugs, sex, alcohol and theft, these results were compared with the Conventional Method of Direct Interview. The study population includes students enrolled in the Semester 2009-II.

KEYWORDS:       RANDOMIZED RESPONSE MODELS  
                      RANDOMIZED MECHANISM  
                      RANDOMIZED RESPONSE MODELS STRATIFIED  
                      DELICATE QUESTION

# ÍNDICE

RESUMEN	vi
SUMMARY	vii
ÍNDICE	viii
INTRODUCCIÓN	x

## **CAPÍTULO I**

### ANTECEDENTES Y CONCEPTOS BÁSICOS

1.1. ANTECEDENTES	1
1.2. CONCEPTOS BÁSICOS	5

## **CAPÍTULO II**

### MODELOS DE RESPUESTA ALEATORIZADA

2.1. MODELOS DE RESPUESTA ALEATORIZADA	7
2.1.1. MODELO DE RESPUESTA ALEATORIZADA DE WARNER	8
2.1.2. MODELO DE LA PREGUNTA NO RELACIONADA O MODELO DE SIMMONS	17
2.1.3. MODELO DE RESPUESTA ALEATORIZADA ESTRATIFICADA DE WARNER	23
2.1.4. MODELO DE RESPUESTA ALEATORIZADA ESTRATIFICADA DE LA PREGUNTA NO RELACIONADA	25

## **CAPÍTULO III**

### APLICACIÓN DEL MODELO DE LA PREGUNTA NO RELACIONADA ESTRATIFICADA – ENCUESTA A ESTUDIANTES SOBRE CONSUMO DE DROGAS, SEXO, CONSUMO DE BEBIDAS ALCOHÓLICAS Y HURTO

3.1. IMPLEMENTACION DEL MODELO DE RESPUESTA ALEATORIZADA ESTRATIFICADA DE LA PREGUNTA NO RELACIONADA	27
3.1.1. ELECCIÓN DE LAS VARIABLES, DEL MECANISMO ALEATORIO Y LA PROBABILIDAD DE SELECCIÓN DE LAS PREGUNTAS	28
3.1.2. TAMAÑO DE MUESTRA	33
3.1.3. PLAN DE MUESTREO DE LA APLICACIÓN	34



3.2.	RESULTADOS DE LA APLICACIÓN PARA EL MODELO DE RESPUESTA ALEATORIZADA ESTRATIFICADA	37
3.2.1.	EVALUACION DEL RENDIMIENTO DE LA MUESTRA	37
3.2.2.	CARACTERÍSTICAS DE LA POBLACIÓN INVESTIGADA	37
3.3.	RESULTADOS DE LA APLICACIÓN PARA EL MÉTODO CONVENCIONAL	38
3.3.1.	EVALUACIÓN DEL RENDIMIENTO DE LA MUESTRA	38
3.3.2.	CARACTERÍSTICAS DE LA POBLACIÓN INVESTIGADA	38
3.4.	COMPARACIÓN DE LOS RESULTADOS NUMÉRICOS DEL MODELO DE RESPUESTA ALEATORIZADA ESTRATIFICADA Y EL MÉTODO CONVENCIONAL	38
3.5.	PRUEBA DE HIPÓTESIS PARA LOS PARÁMETROS EN EL MODELO DE RESPUESTA ALEATORIZADA ESTRATIFICADA Y EL MÉTODO CONVENCIONAL	44
	CONCLUSIONES Y RECOMENDACIONES	50
	BIBLIOGRAFÍA	52
	ANEXOS	54

## INTRODUCCIÓN

En encuestas donde se involucran preguntas altamente delicadas o personales, se ha detectado que, muchas veces las personas no están inclinadas a responder con honestidad o simplemente señalan una negativa por respuesta. Es decir la no respuesta, respuestas evasivas y respuestas falsas, son tan comunes en la práctica, que son difíciles de medir y controlar, para solucionar este problema, es posible utilizar otros procedimientos alternativos, si nuestra intención es obtener datos confiables en esta clase de investigación, en lugar de realizar encuestas por la modalidad de entrevista directa.

Precisamente, el objetivo del presente trabajo es mostrar la utilidad de los Modelos de Respuesta Aleatorizada (MRA), ya que la aplicación de estos modelos disminuye la no respuesta y el sesgo de respuesta en encuestas con preguntas delicadas (Gasto en bebidas alcohólicas, ingresos familiares, número de abortos, etc.).

Este trabajo de investigación comprende de 3 Capítulos, el Capítulo I comprende Antecedentes y Conceptos Básicos para una mejor comprensión del MRA, el Capítulo II contiene el marco teórico de los MRA y por último el Capítulo III presenta una aplicación realizada en la Facultad de Ciencias Matemáticas, sus resultados y una comparación con el Método Convencional (entrevista directa).

## **CAPÍTULO I**

### **ANTECEDENTES Y CONCEPTOS BÁSICOS**

#### **1.1. ANTECEDENTES**

La primera técnica de Respuesta Aleatorizada fue introducida por Warner (1965) con el propósito de obtener datos confiables; para esto se diseña hábilmente las preguntas, de tal manera que, no revela la identidad del entrevistado en el curso de la encuesta. El entrevistador debe utilizar un mecanismo aleatorio (urna con bolas, flecha giratoria, una moneda etc.) para seleccionar una de dos preguntas (Pertenezco al grupo A o No pertenezco al grupo A) cada una de las cuales requiere de una respuesta "SI" o "NO" por parte del entrevistado, sin revelar al entrevistador su posición personal con respecto a la pregunta.

El Modelo estima la proporción de elementos de la población que pertenece al grupo A, a partir de una muestra al azar con reposición.

Simmons (1969), modificó el Modelo original de Warner, con la finalidad de que el entrevistado tenga una mayor disposición a colaborar, sustituyendo el segundo enunciado "No pertenezco al grupo A" por una pregunta referente, a una característica B, no relacionada con A.

El entrevistado debe seleccionar a través de un mecanismo aleatorio, una de las siguientes proposiciones, a las que se supone responde correctamente:

- 1) Pertenezco al Grupo A
- 2) Pertenezco al Grupo B

Los parámetros que se desean estimar son:

$\pi_A$ : Proporción de elementos de la población que pertenecen a A.

$\pi_B$ : Proporción de elementos de la población que pertenecen a B.

Estos dos parámetros se estiman en base a dos muestras aleatorias independientes de tamaño  $n_1$  y  $n_2$  respectivamente.

Greenberg (1971), sugiere una extensión del modelo de la "pregunta no relacionada" para características cuantitativas. En este modelo se utiliza dos variables cuantitativas, la primera variable es  $X$  cuyo valor resulta comprometedor al entrevistado (Gasto en bebidas alcohólicas, número de abortos, ingresos familiares etc.), la segunda variable es,  $Y$  no relacionada a  $X$ ,  $g(\cdot)$  es función de densidad de  $X$  y  $h(\cdot)$  función de densidad de  $Y$ .

El entrevistado utiliza un mecanismo aleatorio, para seleccionar la pregunta delicada con probabilidad  $P$  y la pregunta "no relacionada" con probabilidad  $(1-P)$ .

Los parámetros que se desean estimar son:

$u_x$ : media poblacional de la variable  $X$ , asociada a la pregunta delicada.

$u_y$ : media poblacional de la variable  $Y$ , asociada a la pregunta "no relacionada".

Estos dos parámetros se estiman en base a dos muestras aleatorias independientes de tamaño  $n_1$  y  $n_2$ , y el estimador está relacionado a la variable aleatoria asociada a la respuesta, llamada  $Z$  y que es la mezcla de  $X$  e  $Y$  según  $P$  y  $(1-P)$ .

Utilizando el Modelo de Greenberg se realizó una encuesta, en el Estado de Carolina del Norte en el año 1968, el tema fue sobre abortos y las variables empleadas fueron de tipo cuantitativo. El MRA es aplicado en 5 Áreas Metropolitanas de la Zona Urbana del Estado de Carolina del Norte, de la cual se seleccionó 2 muestras independientes de conglomerados estratificados de 623 y 287 personas.

La población en estudio estuvo comprendida de mujeres mayores de 31 años. El objetivo del estudio fue estimar el promedio de abortos inducidos por raza (pregunta delicada) y el promedio de niños que debería tener una mujer si trabaja a tiempo completo también clasificados por raza (pregunta no relacionada).

En el Colegio de Calcutta en la India, se realizó una encuesta, las variables en estudio fueron de tipo cuantitativo.

Se seleccionó dos muestras el Gasto promedio en bebidas alcohólicas en un determinado mes (pregunta delicada) y el Gasto promedio en libros y útiles que utiliza para estudiar (pregunta no relacionada).

En el año 1982, se aplicó un MRA en una sub-muestra de la Encuesta Epidemiológica de CEDRO. Las variables empleadas fueron de tipo cualitativo.

El MRA de Warner es aplicado en una sub-muestra de Conglomerados de Lima Metropolitana, para investigar la “prevalencia de vida” y el “uso actual” de la sustancia psicoactiva Pasta Básica de Cocaína (PBC). La población en estudio estuvo comprendida con personas de 12 a 50 años de edad.

Liu (1976) nos muestra una descripción de un MRA para variables cuantitativas discretas, en la cual se utiliza un mecanismo aleatorio que es una botella. El recipiente, de nuestra mencionada figura, contiene un número predeterminado de bolas rojas y blancas. En la superficie de las bolas blancas, se les marca con un número  $0, 1, 2, \dots, k$ .

El entrevistado es requerido para voltear el recipiente hacia abajo moviéndolo fuertemente para que una de las bolas aparezcan en la ventana del recipiente. La bola que se asome a la ventana puede ser blanca o roja. Si una bola roja se asoma, el entrevistado puede ser requerido para responder la pregunta delicada (Ejemplo número de abortos que ella tuvo). Si se asoma una bola blanca, el entrevistador leerá simplemente el número.

Los entrevistados se ubican a una distancia considerable para no enterarse del color de la bola que aparece en la ventana del recipiente.

El método aquí presentado, es una modificación del Modelo de Greenberg; es decir, el MRA para variables cuantitativas discretas utiliza un número determinado de bolas blancas, en tanto que el Modelo de Greenberg usa una pregunta no relacionada. La eficiencia del estimador de este modelo depende solamente de la razón del total de bolas y el total de bolas rojas en el recipiente.

Pollock (1976) compara el Modelo de Greenberg con otros MRA para variables cuantitativas. Los cuales son:

- 1) Modelo de Adición.
- 2) Modelo de Multiplicación.

El Modelo de Adición utiliza un método específico para generar  $Y$ , la cual es una “variable aleatoria”, no negativa continua, tiene distribución conocida y está distribuido independiente de la variable cuantitativa  $X$  (asociada a la pregunta delicada).

Luego de obtener  $Y$ , le adiciona a su respuesta asociada a  $X$  y el entrevistador recibe la respuesta:  $Z = X + Y$ .

En el Modelo de Multiplicación: luego de obtener  $Y$ , la multiplica a su respuesta asociada a  $X$  y el entrevistador recibe la respuesta:  $Z = XY$ .

Además, en ambos modelos, se muestra resultados acerca de la estimación de la media y la varianza de  $X$ .

## **1.2. CONCEPTOS BÁSICOS**

### **1.2.1. RESPUESTA ALEATORIZADA Y MODELO DE RESPUESTA ALEATORIZADA**

#### **RESPUESTA ALEATORIZADA**

Es la respuesta a la pregunta de interés, la cual es elegida utilizando un mecanismo aleatorio (urna con bolas, una moneda, flecha giratoria, etc.).

#### **MODELO DE RESPUESTA ALEATORIZADA**

Es la representación matemática de una fórmula, que cumple ciertas condiciones, para dar solución al problema de la no respuesta en encuestas de temas delicados, ejemplo el modelo de Warner, la cual depende de la probabilidad de seleccionar la pregunta delicada, las respuestas de los entrevistados y el tamaño de muestra.

### **1.2.2 PREGUNTA DIRECTA, DELICADA Y NO RELACIONADA**

#### **PREGUNTA DIRECTA**

Cuando el entrevistador hace la pregunta personalmente (de persona a persona).

#### **PREGUNTA DELICADA**

Son aquellas preguntas sensibles o comprometedoras (ingresos, origen de sus rentas, evasión de impuestos, conducta sexual, etc.).

#### **PREGUNTA NO RELACIONADA**

Son aquellas preguntas que no están relacionadas a la pregunta de interés, pero que sin embargo influyen para que la respuesta a la pregunta delicada sea verdadera. Por ejemplo, ¿Es Ud. fumador? (pregunta delicada), ¿Es Ud. lector de algún periódico? (pregunta no relacionada).

### **1.2.3. NO RESPUESTA, TASA Y SESGO DE RESPUESTA**

#### **NO RESPUESTA**

Cuando la persona o entidad que se trata de entrevistar rehúsa su colaboración, (la cual se denomina “negativa”) ya sea por razones personales o subjetivas. El problema es más frecuente cuando el cuestionario hace referencia a datos de conducta íntima o personal (gasto en bebidas alcohólicas, abortos, consumo de drogas, etc.) especialmente si el entrevistado tiene la creencia de que su respuesta puede perjudicarlo.

#### **TASA DE RESPUESTA**

Es el porcentaje que resulta de dividir el total de respuestas obtenidas entre el total de encuestas realizadas en un determinado estudio, multiplicado por 100.

#### **SESGO DE RESPUESTA**

Es la diferencia entre la respuesta esperada y la respuesta obtenida.



## **CAPÍTULO II**

### **MODELOS DE RESPUESTA ALEATORIZADA**

#### **2.1. MODELO DE RESPUESTA ALEATORIZADA**

En encuestas bajo la modalidad de entrevistas directa, generalmente los entrevistados tienden a dar respuestas deliberadamente falsas, respuestas evasivas y no respuestas, cuando se formula preguntas altamente delicadas o personales. En estas llamadas “preguntas comprometedoras”, los entrevistados tienden a dar este tipo de respuesta, ya sea por prejuicios sociales o por no autoinculparse de una conducta.

Una de las alternativas a este problema son los MRA: el Modelo original de Warner y todas sus modificaciones posteriores. Los objetivos de estos diseños son:

- 1) Incrementar la tasa de respuesta.
- 2) Disminuir el sesgo de respuesta.

Los MRA suponen que la actitud del entrevistado se debe a que su reputación puede estar comprometida debido a su posible respuesta, para evitar esto proponen procedimientos que garanticen el anonimato, con el fin de lograr la máxima cooperación de los entrevistados.

Warner, en 1965 utilizó un artificio de aleatorización y mostró que se puede estimar la proporción de personas que pertenecen a determinado grupo, sin que el entrevistado revele su posición personal respecto a la pregunta.

A partir del modelo de Warner se han desarrollado otros modelos para

estimar proporciones y otros parámetros para variables cuantitativas, en situaciones similares.

### **2.1.1. MODELO DE RESPUESTA ALEATORIZADA DE WARNER**

Supongamos que la característica (variable) representada por la letra A, resulta comprometedor para el entrevistado. Una pregunta adecuada a este tipo de variable será: “Pertenece al grupo “A”; si la respuesta es afirmativa, la información corresponderá a este grupo, de lo contrario a su complemento”.

Para este tipo de variables, el Modelo estima la proporción de elementos de la población que pertenece al grupo A, a partir de una muestra simple al azar con reposición de  $n$  personas.

El entrevistador debe utilizar un mecanismo aleatorio (una moneda, urna con bolsa, flecha giratoria, etc.) para seleccionar una de dos preguntas, cada uno de los cuales requiere de una respuesta “SI” o “NO” por parte del entrevistado, sin revelar al entrevistador su posición personal con respecto a la pregunta. Las dos proposiciones a las que se supone responde correctamente son:

- 1) Pertenezco al Grupo A, representada con probabilidad  $p$ .
- 2) No pertenezco al Grupo A, representada con probabilidad  $1-p$ .

La probabilidad de que el mecanismo aleatorio seleccione la proposición (1) “pertenezco al grupo A” es conocida por el muestrista y no por el entrevistador.

Sea:

$n$ : tamaño de la muestra simple al azar con reposición.

$\pi$ : la proporción de elementos de la población que pertenecen al grupo A.

$p$ : la probabilidad de que el mecanismo aleatorio seleccione la proposición “pertenece al grupo A”.

$x_i = 1$ , si el  $i$ -ésimo elemento de la muestra responde SI.

$x_i = 0$ , si el  $i$ -ésimo elemento de la muestra responde NO.

Luego,

$$P = P(x_i = 1) = \pi p + (1-\pi)(1-p) \quad \text{II.1}$$

$$Q = 1 - p = p(x_i = 0) = (1-\pi)p + \pi(1-p) \quad \text{II.2}$$

Se observa que  $x_i$  es una variable aleatoria de Bernoulli, cuya función de densidad es:

$$f_x(x) = [\pi p + (1-\pi)(1-p)]^x [1-\pi]p + \pi(1-p)^{1-x} \quad \text{si } x=0,1$$

$$f_x(x) = 0 \quad , \text{ en otro caso}$$

### ESTIMADOR MÁXIMO VEROSÍMIL DE $\pi$

$$\hat{\pi}_W = \frac{\sum_{i=1}^n x_i}{n(2p-1)} + \frac{(p-1)}{(2p-1)} \quad \text{II.3}$$

el subíndice W indica Estimador de Warner.

Siempre que  $p \neq 1/2$ .

### DEMOSTRACIÓN

La función de verosimilitud es:

$$L(\pi, x_i, p) = [\pi p + (1-\pi)(1-p)]^{\sum x_i} [(1-\pi)p + \pi(1-p)]^{n - \sum x_i}$$

tomando logaritmo,

$$\text{Log} L(\pi, x_i, p) = \sum x_i \text{Log}[\pi p + (1-\pi)(1-p)] + (n - \sum x_i) \text{Log}[(1-\pi)p + \pi(1-p)]$$

$$\frac{\partial \log L(\pi, x_i, p)}{\partial \pi} = 0$$

$$\frac{\sum x_i (2p-1)}{\pi p + (1-\pi)(1-p)} = \frac{(n - \sum x_i)(2p-1)}{(1-\pi)p + \pi(1-p)}$$

$$\sum x_i [(1-\pi)p + \pi(1-p)] = (n - \sum x_i) [\pi p + ((1-\pi)(1-p))]$$

$$\sum x_i [p + \pi - 2\pi p + \pi p + (1-\pi)(1-p)] = n[\pi p + (1-\pi)(1-p)]$$

$$\sum x_i [p + \pi - 2\pi p + \pi p + 1 - p - \pi + \pi p] = n[\pi p + (1-\pi)(1-p)]$$

$$\sum x_i [p + \pi - 2\pi p] = n[\pi p + (1-\pi)(1-p)] - \sum x_i [\pi p + (1-\pi)(1-p)]$$

$$\frac{\sum_{i=1}^n x_i}{n} = \pi p + 1 - p - \pi + \pi p$$

$$\frac{\sum_{i=1}^n x_i}{n} + p - 1 = \pi(2p - 1)$$

Por lo tanto, para  $p \neq 1/2$  se tiene que:

$$\hat{\pi}_w = \frac{\sum_{i=1}^n x_i}{n(2p - 1)} + \frac{(p - 1)}{(2p - 1)}$$

Es el estimador máximo verosímil de  $\pi$ .

**ESPERANZA DEL ESTIMADOR  $\hat{\pi}_w$**

$$E[\hat{\pi}_w] = \frac{E\left[\sum_{i=1}^n x_i\right]}{n(2p - 1)} + \frac{p - 1}{2p - 1}$$

$$E[\hat{\pi}_w] = \frac{1}{2p - 1} \left[ \frac{nE[x_i]}{n} + p - 1 \right] \text{ ya que } E(\sum x_i) = nE(x_i)$$

Considerando (II.1) y (II.2) tenemos:

$$E[\hat{\pi}_w] = \frac{1}{2p-1} [p-1 + \pi p + (1-\pi)(1-p)]$$

$$E[\hat{\pi}_w] = \frac{1}{2p-1} (2\pi p - \pi)$$

$$E[\hat{\pi}_w] = \frac{1}{2p-1} (2p-1)\pi$$

Por lo tanto,  $E[\hat{\pi}_w] = \pi$  y  $\hat{\pi}_w$  es un estimador insesgado de  $\pi$ .

#### **VARIANZA DEL ESTIMADOR $\hat{\pi}_w$**

$$\text{Var}(\hat{\pi}_w) = \frac{\pi(1-\pi)}{n} + \frac{p(1-p)}{n(2p-1)^2} \quad \text{II.4}$$

#### **DEMOSTRACIÓN**

$$\text{Var}(\hat{\pi}_w) = \frac{\text{Var}(\sum x_i)}{(2p-1)^2 n^2}$$

Como  $x_i$  es una variable aleatoria de Bernoulli, su varianza es  $\text{Var}(x_i) = p(1-p)$  y  $\text{Var}(\sum x_i) = n \text{Var}(x_i)$ , de (II.1) y (II.2) tenemos:

$$\text{Var}(\hat{\pi}_w) = \frac{\text{Var}(x_i)}{(2p-1)^2 n}$$

$$Var(\hat{\pi}_w) = \frac{[\pi p + (1-\pi)(1-p)][(1-\pi)p + \pi(1-p)]}{(2p-1)^2 n}$$

$$Var(\hat{\pi}_w) = \frac{\pi p^2(1-\pi) + \pi p^2(1-p) + (1-\pi)^2 p(1-p) + \pi(1-\pi)(1-p)^2}{(2p-1)^2 n}$$

$$Var(\hat{\pi}_w) = \frac{p(1-p)[\pi^2 + (1-\pi)^2] + \pi(1-\pi)[p^2 + (1-p)^2]}{(2p-1)^2 n}$$

$$Var(\hat{\pi}_w) = \frac{p(p-1)(\pi^2 - \pi + 1/2) + \pi(1-\pi)(p^2 - p + 1/2)}{1/2(2p-1)^2 n}$$

Completando cuadrados, se tiene:

$$Var(\hat{\pi}_w) = \frac{(p-p^2)[(\pi-1/2)^2 + 1/4] + (\pi-\pi^2)[(p-1/2)^2 + 1/4]}{1/2(2p-1)^2 n}$$

$$Var(\hat{\pi}_w) = \frac{1}{2(2p-1)^2 n} \left\{ -(p^2 - p + 1/4 - 1/4)[(\pi-1/2)^2 + 1/4] + [-(\pi^2 - \pi + 1/4 - 1/4)][(p-1/2)^2 + 1/4] \right\}$$

$$Var(\hat{\pi}_w) = \frac{1}{2(2p-1)^2 n} \left\{ -[(p-1)^2 - 1/4][(\pi-1/2)^2 + 1/4] + [(\pi-1/2)^2 - 1/4][(p-1/2)^2 + 1/4] \right\}$$

Resolviendo tenemos:

$$Var(\hat{\pi}_w) = \frac{-2(p-1/2)^2(\pi-1/2)^2 + 2/16}{1/2(2p-1)^2 n}$$

$$Var(\hat{\pi}_w) = \frac{2[1/16 - (p-1/2)^2(\pi-1/2)^2]}{1/2(2p-1)^2 n}$$

$$Var(\hat{\pi}_w) = \frac{2[1/16 - (p-1/2)^2(\pi-1/2)^2]}{2(p-1/2)^2 n}$$

$$Var(\hat{\pi}_w) = \frac{1}{n} \left[ \frac{1}{16(p-1/2)^2} - (\pi-1/2)^2 \right]$$

$$Var(\hat{\pi}_w) = \frac{1}{n} \left[ \frac{1}{16(p-1/2)^2} - 1/4 + 1/4 - (\pi-1/2)^2 \right]$$

$$Var(\hat{\pi}_w) = \frac{1}{n} \left[ \frac{4-16(p-1/2)^2}{4 \times 16(p-1/2)^2} + 1/4 - \pi^2 + \pi - 1/4 \right]$$

$$Var(\hat{\pi}_w) = \frac{1}{n} \left[ \frac{4-16p^2+16p-4}{4 \times 16(p-1/2)^2} + \pi(1-\pi) \right]$$

$$Var(\hat{\pi}_w) = \frac{\pi(1-\pi)}{n} + \frac{p(1-p)}{n(2p-1)^2}$$

Es la varianza del estimador  $\hat{\pi}_w$

Observamos que, la varianza del estimador de Warner tiene dos componentes, el primer sumando es la varianza de la proposición muestral bajo muestreo aleatorio al azar con reposición y el segundo es la parte de la varianza del procedimiento aleatorio. Es decir, que el uso del procedimiento de Warner incrementa la varianza en este último término.

## ELECCIÓN DE P Y DEL TAMAÑO DE MUESTRA

Existen dos preguntas importantes que debemos hacer:

### 1. ¿Cómo será la cooperación de las personas entrevistadas?

Warner observa que la cooperación del entrevistado está en función del grado de “anonimato” y dependerá de p.

$$\hat{\pi}_w = \frac{\sum_{i=1}^n x_i}{n(2p-1)} + \frac{(p-1)}{(2p-1)}$$

Si  $p=1/2$ , el estimador de Warner (II.3) no está definido y por lo tanto, la muestra no arroja información sobre el parámetro.

Si  $p=1$ , el estimador de Warner, se reduce a la estimación convencional de  $\pi_A$ , que requiere que el entrevistado informe sin reservas si pertenece o no al grupo A.

Si  $p=0$ , el estimador de Warner, se reduce a la estimación convencional de  $(1-\pi_A)$ .

Para elegir “p”, veremos en que intervalos de éste, la varianza del estimador se hace mínima y cuando máxima.

### **VARIANZA DEL ESTIMADOR DE $\hat{\pi}_w$ COMO FUNCIÓN MONÓTONA CRECIENTE EN EL INTERVALO $p \in [0, 1/2]$**

$$Var(\hat{\pi}_w) = \frac{1}{n} \left[ \pi(1-\pi) + \frac{p(1-p)}{(2p-1)^2} \right] \quad \text{para } p \neq 1/2 \text{ y } 0 \leq p \leq 1$$

Derivado parcialmente con respecto a p, tenemos:

$$\frac{\partial Var(\hat{\pi}_w)}{\partial p} = \frac{1}{n} \left[ \frac{(1-2p)(2p-1)^2 - p(1-p)2(2p-1)(2)}{(2p-1)^4} \right]$$

$$\frac{\partial Var(\hat{\pi}_w)}{\partial p} = \frac{1}{n} \left[ \frac{(1-2p)(2p-1) - 4p(1-p)}{(2p-1)^3} \right]$$

$$\frac{\partial Var(\hat{\pi}_w)}{\partial p} = \frac{1}{n} \left[ \frac{-(1-2p)^2 - 4p + 4p^2}{(2p-1)^3} \right]$$



$$\frac{\partial Var(\hat{\pi}_w)}{\partial p} = -\frac{1}{n(2p-1)^3}$$

$$\frac{\partial Var(\hat{\pi}_w)}{\partial p} = -\frac{1}{n(2p-1)^3} > 0 \quad \forall p \in [0, 1/2) \text{ y } n \in \mathbb{N}$$

Para  $n$  y  $\pi$  fijos, la  $Var(\hat{\pi}_w)$  es una función creciente en  $p$  en el intervalo  $[0, 1/2)$ .

**VARIANZA DEL ESTIMADOR DE  $\hat{\pi}_w$  COMO FUNCIÓN MONÓTONA DECRECIENTE EN EL INTERVALO  $p \in (1/2, 1]$ .**

$$\frac{\partial Var(\hat{\pi}_w)}{\partial p} = -\frac{1}{n(2p-1)^3} < 0 \quad \forall p \in (1/2, 1] \text{ y } n \in \mathbb{N}$$

Para  $n$  y  $\pi$  fijos, la  $Var(\hat{\pi}_w)$  es una función decreciente para  $p$  en el intervalo  $(1/2, 1]$ .

Por lo tanto, la elección de  $p$ , depende de la varianza del estimador de Warner. Para  $p$  cercano a 1 ó cercano a 0 la varianza se hace pequeña y para un  $p$  cercano a  $1/2$  la varianza comienza a crecer indefinidamente.

## 2. ¿Qué tamaño de muestra es el requerido para obtener varios niveles de precisión del estimador?

Para esto se diseña el Cuadro N°(II.1) que se exhibe más adelante, para lo cual determinaremos el tamaño de muestra de la siguiente manera:

Sabemos que,

$$C.V.(\hat{\pi}_w) = \frac{\sqrt{Var(\hat{\pi}_w)}}{\pi}$$

Despejando la varianza del estimador  $\pi_w$ .

$$Var(\hat{\pi}_w) = \pi^2 [C.V.(\hat{\pi})^2]$$

Reemplazando el resultado anterior, en (II.4) tenemos:

$$[\pi(C.V.(\hat{\pi}))]^2 = \frac{\pi(1-\pi)}{n} + \frac{p(1-p)}{n(2p-1)^2}$$

Despejando n:

$$[\pi(C.V.(\hat{\pi}))]^2 = \frac{1}{n} \left[ \pi(1-\pi) + \frac{p(1-p)}{(2p-1)^2} \right]$$

$$n = \frac{\pi(1-\pi) + \frac{p(1-p)}{(2p-1)^2}}{[\pi(C.V.(\hat{\pi}_w))]^2}$$

Para distintos valores de  $\pi$ , C.V.  $(\hat{\pi}_w)$  y p se obtienen los tamaños de muestra que se presentan en el cuadro mencionado.

Fijando  $\pi$  y el C.V.  $(\hat{\pi}_w)$  al crecer p el tamaño de la muestra disminuye, si se tienen p cercano a 1 se requiere de una muestra más pequeña, que la requerida para p cercana a 1/2.

Cuadro N° II.1 Tamaño de muestra

$\pi$	C.V. $(\hat{\pi}_w)$	MRA					n convencional
		p=.70	p=.75	p=.80	p=.85	p=.90	
0.15	0.05	25600	15600	10167	6892	4770	2266
0.20	0.05	14725	9100	6040	4202	3006	1600
0.30	0.05	6767	4267	2907	2090	1558	933
0.40	0.05	3881	2475	1710	1250	952	600
0.50	0.05	2500	1600	1110	816	625	400
0.50	0.10	625	400	278	204	156	100
0.40	0.10	970	619	428	313	238	150
0.30	0.10	1692	1067	727	522	390	233
0.20	0.10	3681	2275	1510	1050	752	400
0.15	0.10	6400	3900	2540	1723	1192	567
0.10	0.10	14025	8400	5344	3502	2306	900
0.05	0.10	54400	31900	19678	12308	7525	1900

De manera similar se observa que, el MRA de Warner incrementa el tamaño de la muestra, en comparación al muestreo simple aleatorio con reposición.

### 2.1.2. MODELO DE LA PREGUNTA “NO RELACIONADA” O MODELO SE SIMMONS

Greenberg, Abul-Ela, Simmons y Horvitz (1969) modificaron el Modelo original de Warner, con el propósito de que el entrevistado tenga una mayor disposición a colaborar, sustituyendo la segunda proposición: “No pertenezco al Grupo A”, por otra pregunta referente a una característica B, no relacionada con A.

En consecuencia el entrevistador debe seleccionar a través de un procedimiento aleatorio (flecha giratoria, una moneda, urna con bolas, etc.) una de las siguientes proposiciones a las que se supone responde correctamente:

- 1) Pertenezco al Grupo A.
- 2) Pertenezco al Grupo B.

Los parámetros que se desean estimar son:

$\pi_A$ : Proporción de elementos de la población que pertenecen a A.

$\pi_B$ : Proporción de elementos de la población que pertenecen a B.

Estos dos parámetros se estiman en base a dos muestras aleatorias independientes de tamaño  $n_1$  y  $n_2$  respectivamente. Para hallar los estimadores, se realiza el desarrollo siguiente:

Cuadro N° II.2: Probabilidad de selección

ENUNCIADO	Prob. de selección		Var. Aleatoria	
	Muestra $n_1$	Muestra $n_2$	$n_1$	$n_2$
1. Pertenezco a A	$p_1$	$p_2$	$X_{1i}$	$X_{2j}$
2. Pertenezco a B	$1-p_1$	$1-p_2$		

$X_{1i} = 1$ , si i-ésimo elemento de la muestra 1 responde SI.

$X_{1i} = 0$ , si i-ésimo elemento de la muestra 1 responde NO.

$X_{2j} = 1$ , si j-ésimo elemento de la muestra 2 responde SI.

$X_{2j} = 0$ , si j-ésimo elemento de la muestra 2 responde NO.

Entonces:

$$P(x_{1i}=1) = p_1\pi_A + (1-p_1)\pi_B, i = 1, 2, \dots, n_1$$

$$P(x_{2j}=1) = p_2\pi_A + (1-p_2)\pi_B, j = 1, 2, \dots, n_2$$

despejando  $\pi_A$  y  $\pi_B$

$$p_1\pi_A = P(x_{1i}=1) - (1-p_1)\pi_B \quad \text{II.5}$$

$$p_2\pi_A = P(x_{2j}=1) - (1-p_2)\pi_B \quad \text{II.6}$$

de (II.5) y (II.6) tenemos:

$$\pi_A = \frac{P(x_{1i}=1) - (1-p_1)\pi_B}{p_1}$$

$$\pi_A = \frac{P(x_{2j}=1) - (1-p_2)\pi_B}{p_2}$$

Igualando se tiene:

$$\frac{P(x_{1i}=1) - (1-p_1)\pi_B}{p_1} = \frac{P(x_{2j}=1) - (1-p_2)\pi_B}{p_2}$$

Luego:

$$\pi_B = \frac{P(x_{1i}=1)p_2 - P(x_{2j}=1)p_1}{p_2 - p_1} \quad \text{II.7}$$

$$\pi_A = \frac{P(x_{1i}=1)(1-p_2) - P(x_{2j}=1)(1-p_1)}{p_1 - p_2} \quad \text{II.8}$$

Como  $x_{1i}$  y  $x_{2j}$  son variables aleatorias de Bernoulli se tiene:

$$E[x_{1i}] = P(x_{1i}=1) = \mu_1 \quad \forall i = 1, 2, \dots, n_1$$

$$E[x_{2j}] = P(x_{2j}=1) = \mu_2 \quad \forall j = 1, 2, \dots, n_2$$

Utilizando el método de los momentos y sustituyendo, tenemos lo siguiente:

$$P(x_{1i}=1) = E[x_{1i}] = \sum_{i=1}^{n_1} \frac{x_{1i}}{n_1} = \hat{\mu}_1$$

$$P(x_{2j}=1) = E[x_{2j}] = \sum_{j=1}^{n_2} \frac{x_{2j}}{n_2} = \hat{\mu}_2$$

De las ecuaciones (II.7) y (II.8), obtenemos las siguientes estimaciones:

$$\hat{\pi}_{A(G)} = \frac{(1-p_2) \frac{\sum_{i=1}^{n_1} x_{1i}}{n_1} + (1-p_1) \frac{\sum_{j=1}^{n_2} x_{2j}}{n_2}}{p_1 - p_2} \quad p_1 \neq p_2$$

$$\hat{\pi}_{B(G)} = \frac{p_2 \frac{\sum_{i=1}^{n_1} x_{1i}}{n_1} + p_1 \frac{\sum_{j=1}^{n_2} x_{2j}}{n_2}}{p_1 - p_2} \quad p_1 \neq p_2$$

**ESPERANZA DE  $\hat{\pi}_{A(G)}$ ,  $\hat{\pi}_{B(G)}$**

$$E[\hat{\pi}_{A(G)}] = \frac{(1-p_2)E[x_{1i}] + (1-p_1)E[x_{2j}]}{p_1 - p_2}$$

$$E[\hat{\pi}_{A(G)}] = \frac{(1-p_2)P(x_{1i}=1) + (1-p_1)P(x_{2j}=1)}{p_1 - p_2}$$

Luego de (II.8) se tiene:

$$E[\hat{\pi}_{A(G)}] = \pi_A \text{ y } \hat{\pi}_{A(G)} \text{ es un estimador insesgado de } \pi_A.$$

En forma similar tenemos que:

$$E[\hat{\pi}_{B(G)}] = \pi_B \text{ y } \hat{\pi}_{B(G)} \text{ es un estimador insesgado de } \pi_B.$$

**VARIANZA DE  $\hat{\pi}_{A(G)}$ ,  $\hat{\pi}_{B(G)}$**

Sabemos que:

$$\hat{\pi}_{(AG)} = \frac{(1-p_2) \frac{\sum_{i=1}^{n_1} x_{1i}}{n_1} + (1-p_1) \frac{\sum_{j=1}^{n_2} x_{2j}}{n_2}}{p_1 - p_2} \quad p_1 \neq p_2$$

Tomando varianza a ambos miembros tenemos:

$$Var[\hat{\pi}_{A(G)}] = \frac{(1-p_2)^2 \frac{\sum_{i=1}^{n_1} V(x_{1i})}{n_1^2} + (1-p_1)^2 \frac{\sum_{j=1}^{n_2} v(x_{2j})}{n_2^2}}{(p_1 - p_2)^2} \quad p_1 \neq p_2$$

$$Var[\hat{\pi}_{A(G)}] = \frac{(1-p_2)^2 \frac{n_1 V(x_{1i})}{n_1^2} + (1-p_1)^2 \frac{n_2 v(x_{2j})}{n_2^2}}{(p_1 - p_2)^2} \quad p_1 \neq p_2$$

Ya que,  $x_{1i}$  y  $x_{2j}$  son variables aleatorias de Bernoulli y  $E[x_{1i}] = P(x_{1i}=1) = \mu_1$

$$Var[\hat{\pi}_{A(G)}] = \frac{(1-p_2)^2 \frac{n_1 u_1 (1-u_1)}{n_1^2} + (1-p_1)^2 \frac{n_2 u_2 (1-u_2)}{n_2^2}}{(p_1 - p_2)^2}$$

Simplificando, se obtiene:

$$Var[\hat{\pi}_{A(G)}] = \frac{(1-p_2)^2 \frac{u_1(1-u_1)}{n_1} + (1-p_1)^2 \frac{u_2(1-u_2)}{n_2}}{(p_1 - p_2)^2}$$

$$Var[\hat{\pi}_{A(G)}] = [(1-p_2)^2 \frac{u_1(1-u_1)}{n_1} + (1-p_1)^2 \frac{u_2(1-u_2)}{n_2}] \frac{1}{(p_1 - p_2)^2}$$

En forma similar se obtiene Varianza del Estimador  $\pi_B$  y utilizando:  $x_{2j}$  es variable aleatoria de Bernoulli y  $E[x_{2j}] = P(x_{2j}=1) = \mu_2$ .

$$\hat{\pi}_{B(G)} = \frac{p_2 \frac{\sum_{i=1}^{n_1} x_{1i}}{n_1} + p_1 \frac{\sum_{j=1}^{n_2} x_{2j}}{n_2}}{p_2 - p_1} \quad p_1 \neq p_2$$

$$Var[\hat{\pi}_{B(G)}] = [p_2^2 \frac{u_1(1-u_1)}{n_1} + p_1^2 \frac{u_2(1-u_2)}{n_2}] \frac{1}{(p_1 - p_2)^2}$$

## EL CASO CUANDO $\pi_B$ ES CONOCIDO

En este caso solo se requiere de una muestra aleatoria, y las proposiciones que se supone correctas son:

- 1) Pertenezco al grupo A, presentada con probabilidad p.
- 2) Pertenezco al grupo B, presentada con probabilidad 1-p.

Donde:

$$X_i = \begin{cases} 1, & \text{si i-ésimo elemento de la muestra responde "SI".} \\ 0, & \text{si i-ésimo elemento de la muestra responde "NO".} \end{cases}$$

$$P(X_i=1) = p\pi_A + (1-p)\pi_B$$

y utilizando el Método de Momentos se tiene:

$$E[x_i=1]=P(x_i=1)=\frac{\sum_i^n x_i}{n}$$

Despejando  $\pi_{A(G)}$ :

Y reemplazando  $P(x_i=1)$  se tiene:

$$\hat{\pi}_{A(G)} = \frac{P(x_i = 1) - (1 - p)\pi_B}{p}$$

$$\hat{\pi}_{A(G)} = \frac{\sum_i^n \frac{x_i}{n} - (1 - p)\pi_B}{p}$$

## ESPERANZA Y VARIANZA

$$E[\hat{\pi}_{A(G)}] = \frac{E[x_i] - (1 - p)\pi_B}{p}$$

$$E[\hat{\pi}_{A(G)}] = \frac{p\pi_A + (1 - p)\pi_B - (1 - p)\pi_B}{p} = \pi_A$$

$\hat{\pi}_{A(G)}$  es un estimador insesgado de  $\pi_A$ .

$$Var[\hat{\pi}_{A(G)}] = \frac{\mu(1 - \mu)}{np^2}$$

Donde:

$$\mu = p\pi_A + (1 - p)\pi_B$$

## ELECCIÓN DE $P_1$ Y $P_2$ , $\pi_B$ y $n$

En el trabajo de Greenberg, Abul-Ela, Simmons y Horvits (1969) se presentan las siguientes consideraciones:

- 1) Para elegir  $p_1$  y  $p_2$ , se debe tener en cuenta:
  - Disminución de la varianza.



- Disposición a colaborar por parte del entrevistado.

Para mantener esta disposición proponen que  $p_1 + p_2 = 1$

- 2) Para elegir la proposición no relacionada  $\pi_B$

Se debe considerar lo siguiente:

$\pi_A < 0.5$  entonces escoger  $\pi_B$  pequeño

$\pi_A > 0.5$  entonces escoger  $\pi_B$  grande

- 3) Afijación óptima de  $n_1$  y  $n_2$  para minimizar la varianza de:

$\hat{\pi}_{A(G)}$  sujeto a la condición  $n_1 + n_2 = n$ .

Esta afijación viene dada por:

$$\frac{n_1}{n_2} = \frac{\mu_1(1 - \mu_1)(1 - p_2)^2}{\mu_2(1 - \mu_2)(1 - p_1)^2}$$

Moors (1971) demuestra que si en el modelo de Simmons se escoge  $p_2=0$ , se obtiene un diseño más eficiente que el obtenido a través de la elección de  $p_2 = 1 - p_2$  y  $n_1$  y  $n_2$  con afijación óptima.

Lanke (1975) muestra que, en muchos casos resulta conveniente escoger la característica B de tal forma que  $\pi_B$  tome valores “grandes”.

Además, considera la relación entre  $p_1$  y  $\pi_B$ , y muestra que: cuanto menor sea  $\pi_B$  menor será  $p_1$  que un entrevistado, esté dispuesto a aceptar para cooperar con la investigación.

### 2.1.3. EL MODELOS DE RESPUESTA ALEATORIZADA ESTRATIFICADA DE WARNER

En el Muestreo Aleatorizado Estratificado la población a investigar se divide en estratos. Una muestra es entonces seleccionada por muestreo aleatorio simple con reemplazamiento de cada estrato para conseguir todo el beneficio de estratificación, se asume que el número de unidades en cada estrato es conocido.

En los Modelos de Warner Estratificado y el MRA de la pregunta no relacionada, un respondiente individual en la muestra del estrato  $i$  es instruido para usar la aleatorización del mecanismo  $R_i$ , el cual consiste de una tarjeta con pregunta sensible ( $S$ ) con probabilidad  $P_i$  y una tarjeta con pregunta negativa ( $\bar{S}$ ) con probabilidad  $1-P_i$ . El respondiente o entrevistado responde la pregunta con un “SI” o “NO” sin reportar que tarjeta ha elegido y por ende que pregunta está respondiendo. Un respondiente de diferente estrato llevará a cabo la asignación al azar de diferentes mecanismos aleatorizados, el cual tiene diferentes probabilidades preasignadas. Bajo la suposición que las respuestas reportadas “SI” o “NO” son verdaderas y  $P_i$  son fijadas por el investigador, la probabilidad de responder “SI” en el estrato “ $i$ ” por el MRA de Warner Estratificado es:

$$Z_i = p_i \pi_{si} + (1 - p_i) (1 - \pi_{si}), \quad \text{para } i = 1, 2, \dots, k$$

Donde:

$Z_i$  : es la proporción de respuesta “SI” en el estrato  $i$

$P_i$  : Probabilidad de seleccionar la pregunta sensible en el estrato  $i$

$\pi_{s_i}$  : es la proporción de respondientes en la población con la pregunta sensible en el estrato  $i$

$n_i$  : denota el número de unidades en la muestra de estratos  $i$ ,  
 $n$  : número total de unidades en las muestras de todos los estratos

### **ESTIMADOR DE $\pi_{s_i}$**

El estimador de  $\pi_{s_i}$  es expresado de la siguiente manera:

$$\hat{\pi}_{s_i} = \frac{Z_i}{2(P_i - 1)} + \frac{(P_i - 1)}{(2P_i - 1)},$$

donde 
$$Z_i = \frac{\sum_{i=1}^n X_i}{n_i}$$

Sea  $n_i$  que denota el número de unidades en las muestras de cada estrato entonces  $n = \sum_{i=1}^k n_i$ , Kim y Warde (2003) muestra la afijación óptima de  $n$  para  $n_1, n_2, \dots, n_{k-1}$  y  $n_k$  para derivar la mínima varianza de  $\pi_s$ , el cual es el siguiente:

$$\frac{n_i}{n} = \frac{w_i \left[ \pi_{s_i} (1 - \pi_{s_i}) + \frac{P_i(1 - P_i)}{(2P_i - 1)^2} \right]^{1/2}}{\sum_{i=1}^k w_i \left[ \pi_{s_i} (1 - \pi_{s_i}) + \frac{P_i(1 - p_i)}{(2P_i - 1)^2} \right]^{1/2}}$$

Donde  $w_i = \frac{N_i}{N}$  para  $i = 1, 2, \dots, k$  ( $N$  es el tamaño de la población y  $N_i$  es el total de elementos o unidades en el estrato  $i$ ). Bajo la suposición de que  $n_i = n \left( \frac{N_i}{N} \right)$  y  $P_i = P$  para todo  $i$ , Kim y Warde (2003) muestra la mínima varianza de  $\hat{\pi}_s$  como sigue:

$$Var(\hat{\pi}_s) = \frac{1}{n} \left[ \sum_{i=1}^k w_i \left\{ \pi_{s_i} (1 - \pi_{s_i}) + \frac{2(1 - p)}{(2p - 1)^2} \right\}^{1/2} \right]^2$$

#### 2.1.4. MODELO DE RESPUESTA ALEATORIZADA ESTRATIFICADA DE LA PREGUNTA NO RELACIONADA

Para el MRA Estratificada de la pregunta no relacionada, la probabilidad de obtener una respuesta “SI” en el estrato “i” es :

$$Z_i = P_i \pi_{s_i} + (1 - p_i) \pi_N, \quad \text{para } i = 1, 2, \dots, k, \quad \text{II.9}$$

donde  $\pi_N$  : se asume conocido y es la proporción de respondientes con la pregunta no sensible en el estrato  $i$ .

### El Estimador de $\pi_{s_i}$

El estimador de  $\pi_{s_i}$  en el Modelo de la pregunta no relacionada es expresado de la siguiente manera:

$$\hat{\pi}_{s_i} = \frac{Z_i - (1 - P_i)\pi_N}{P_i}, \quad \text{donde } Z_i = \frac{\sum_{i=1}^n X_i}{n_i}. \quad \text{II.10}$$

Kim y Elam (2003a) muestra la afijación óptima de  $n$  para  $n_1, n_2, \dots, n_{k-1}$  y  $n_k$ , para derivar la mínima varianza de  $\pi_s$ , el cual es el siguiente:

$$\frac{n_i}{n} = \frac{\frac{w_i}{p_i} \sqrt{z_i(1-z_i)}}{\sum_{i=1}^k \frac{w_i}{p_i} \sqrt{z_i(1-z_i)}} \quad \text{II.11}$$

Kim y Elam (2003a) también da la varianza mínima de  $\hat{\pi}_s$  como sigue:

$$\text{var}(\hat{\pi}_s) = \frac{1}{n} \left[ \sum_{i=1}^k \frac{w_i}{p_i} \sqrt{z_i(1-z_i)} \right]^2 \quad \text{II.12}$$

**CAPITULO III**  
**APLICACIÓN DEL MODELO DE LA PREGUNTA NO RELACIONADA**  
**ESTRATIFICADA – ENCUESTA A ESTUDIANTES SOBRE CONSUMO DE**  
**DROGAS, SEXO, CONSUMO DE BEBIDAS ALCOHÓLICAS Y HURTO**

**3.1. IMPLEMENTACION DEL MODELO DE RESPUESTA ALEATORIZADA**  
**ESTRATIFICADA DE LA PREGUNTA NO RELACIONADA**

EL MRA de la pregunta no relacionada estratificada para características cualitativas (Kim and Elam (2003a)) será utilizado en una encuesta sobre prevalencia y consumo actual de drogas, sexo, consumo de alcohol y hurto. Su aplicación se hará en la Facultad de Ciencias Matemáticas de la Universidad Nacional Mayor de San Marcos, utilizando el esquema de muestreo estratificado, donde cada estrato es una Escuela Académico Profesional, seleccionados al azar del listado de Alumnos Matriculados en el Semestre 2009-II.

El propósito de esta investigación es mostrar la utilidad de la técnica de Respuesta Aleatorizada (RA), para obtener una estimación de algunas variables como la prevalencia de vida de drogas, proporción de alumnos que consumen actualmente drogas, proporción de alumnos que consumen bebidas alcohólicas con frecuencia etc., reduciendo principalmente el sesgo de respuesta (por respuestas deliberadamente falsas) y obtener tasas de respuestas mayores.

### 3.1.1. ELECCIÓN DE LAS VARIABLES, ELECCIÓN DEL MECANISMO ALEATORIO Y DE LA PROBABILIDAD DE SELECCIONAR LAS PREGUNTAS

Las variables elegidas para este estudio serán las siguientes:

#### 1. PREVALENCIA DE VIDA DE DROGAS

Diez tarjetas de color Blanco

- Ocho con el enunciado:  
A) “Durante mi vida he consumido, pasta básica de cocaína (pasta, pay o pastel) por lo menos una vez”
- Dos con el enunciado:  
B) ¿Utiliza Messenger?

Luego la probabilidad (P) de seleccionar el enunciado A) es 0,80 y la probabilidad (1-P) de seleccionar el enunciado B) es 0,20.

Figura III.1

Modelo de tarjetas: “Prevalencia de vida de drogas”

P = 0,80

Durante mi vida he consumido, pasta básica de cocaína (pasta, pay o pastel) por lo menos una vez

Lea en silencio y responda en voz alta

Sí

O

No

1-P = 0,20

¿Utiliza ud. Messenger?

Lea en silencio y responda en voz alta

Sí

O

No

## 2. USO ACTUAL DE DE DROGAS

Diez tarjetas de color Palo Rosa

- Ocho con el enunciado:  
A) “Actualmente soy consumidor de pasta básica de cocaína (pasta, pay o pastel)”
- Dos con el enunciado;  
B) ¿Utiliza ud. Página Web?

Luego la probabilidad (P) de seleccionar el enunciado A) es 0,80 y la probabilidad (1-P) de seleccionar el enunciado B) es 0,20.

Figura III.2

Modelo de tarjetas: “Uso actual de drogas”

P =0,80

Actualmente soy consumidor de pasta básica de cocaína (pasta, pay o pastel)		
Lea en silencio y responda en voz alta		
Sí	O	No

1-P=0,20

¿Utiliza ud. página Web?		
Lea en silencio y responda en voz alta		
Sí	O	No

### 3. PREVALENCIA DE SEXO

Diez tarjetas de color Rosado

- Ocho con el enunciado:  
A) ¿Ha mantenido relaciones sexuales con más de dos personas a lo largo de su vida?
- Dos con el enunciado;  
B) ¿Alguna vez interrumpió sus estudios universitarios?

Luego la probabilidad (P) de seleccionar el enunciado A) es 0,80 y la probabilidad (1-P) de seleccionar el enunciado B) es 0,20.

Figura III.3

Modelo de tarjetas : “Prevalencia de sexo”

P =0,80

¿Ha mantenido relaciones sexuales con más de dos personas a lo largo de su vida?

Lea en silencio y responda en voz alta

Si

O

No

1- P=0,20

¿Alguna vez interrumpió sus estudios universitarios?

Lea en silencio y responda en voz alta

Sí

O

No



#### 4. CONSUMO ACTUAL DE ALCOHOL

Diez tarjetas de color Celeste

- Ocho con el enunciado:  
A) ¿Consume alcohol (cerveza, vino, sangría, etc) con frecuencia? (todos los fines de semana)
- Dos con el enunciado;  
B) ¿Trabaja ud. actualmente?

Luego la probabilidad (P) de seleccionar el enunciado A) es 0,80 y la probabilidad (1-P) de seleccionar el enunciado B) es 0,20.

Figura III.4

Modelo de tarjetas: "Consumo actual de alcohol"

P =0,80

¿Consume alcohol (cerveza, vino, sangría, etc) con frecuencia? (todos los fines de semana)		
Lea en silencio y responda en voz alta		
Si	O	No

1 - P=0,20

¿Trabaja ud. actualmente?		
Lea en silencio y responda en voz alta		
Sí	O	No

## 5. PREVALENCIA DE HURTO

Diez tarjetas de color Naranja

- Ocho con el enunciado:  
A) He llevado o he consumido sin pagar algún producto de algún supermercado
- Dos con el enunciado;  
B) ¿Ud. eligió su carrera profesional por vocación?

Luego la probabilidad (P) de seleccionar el enunciado A) es 0,80 y la probabilidad (1-P) de seleccionar el enunciado B) es 0,20.

Figura III.5

Modelo de tarjetas: "Prevalencia de hurto"

P =0,80

He llevado o he consumido sin pagar algún producto de algún supermercado

Lea en silencio y responda en voz alta

Si

O

No

1- P=0,20

¿Ud. eligió su carrera profesional por vocación?

Lea en silencio y responda en voz alta

Sí

O

No

### 3.1.2. TAMAÑO DE MUESTRA

De acuerdo a Registros Académicos de la Dirección Académica los Alumnos Matriculados el Semestre 2009 - II en la Facultad de Ciencias Matemáticas de la UNMSM, está dividida en cuatro Escuelas académico Profesionales (Ver Cuadro N° III.1), de acuerdo a esta información se decidió considerar a cada Escuela Académico Profesional como un estrato, en total tenemos 4 estratos o escuelas.

Cuadro N° III.1

Distribución de los Estudiantes de Pre-Grado según Escuela Académico Profesional – Semestre 2009-II - FCM - UNMSM

<b>Escuela Académico Profesional</b>	<b>Población</b>	<b>%</b>
Matemática	423	29.33
Estadística	271	18.79
Investigación Operativa	457	31.69
Computación Científica	291	20.18
<b>Total</b>	<b>1442</b>	<b>100.00</b>

*Fuente: Dirección Académica - FCM – UNMSM*

El esquema de muestreo que se utilizó fue el Muestreo Aleatorio Estratificado (MAE), con afijación proporcional de acuerdo a la cantidad de alumnos en cada uno de los estratos, donde cada estrato es una Escuela Académico Profesional de la Facultad de Ciencias Matemáticas de la Universidad Nacional Mayor de San Marcos.

Se utilizó el MAE (Cochran, 1977), con un nivel de confianza del 95% y un margen de error del 4,5%, el tamaño de muestra fue de 301 alumnos matriculados el Semestre 2009-II. El MRA requiere un tamaño de muestra más grande que el Método Convencional. La distribución de la muestra por Escuela Académico Profesional, para ambos métodos, se muestra en el Cuadro III.2.

Cuadro N° III.2

Distribución de la muestra para ambos Métodos según Escuela Académico Profesional (Semestre 2009-II), FCM – UNMSM

<b>Escuela Académico Profesional</b>	<b>Muestra MRA</b>	<b>Muestra M. Convenc.</b>
Matemática	88	60
Estadística	57	40
Investigación Operativa	95	64
Computación Científica	61	36
<b>Total</b>	<b>301</b>	<b>200</b>

### 3.1.3. PLAN DE MUESTREO DE LA APLICACIÓN

#### 1) POBLACIÓN OBJETIVO

La población objetivo, es aquella acerca de la cual deseamos recolectar la información, en nuestro caso, es la población de alumnos matriculados el Semestre 2009-II, en la Facultad de Ciencias Matemáticas, de la Universidad Nacional Mayor de San Marcos.

#### 2) MARCO MUESTRAL

El Marco muestral esta constituido por el listado de Alumnos Matriculados el Semestre Académico 2009-II, en la Facultad de Ciencias Matemáticas de la Universidad Nacional Mayor de San Marcos, la cual fue proporcionado por Secretaria Académica de dicha Facultad.

#### 3) DISEÑO DE LA MUESTRA

Se dividió el marco muestral en 4 estratos:

El estrato 1 lo conforma la EAP de Matemática.

El estrato 2 lo conforma la EAP de Estadística

El estrato 3 lo conforma la EAP de Investigación Operativa.

El estrato 4 lo conforma la EAP de Computación Científica.

Para elegir las unidades en el estrato 1, se enumeró el listado de la EAP de Matemática desde 1 hasta 423. Y utilizando el siguiente procedimiento en el Programa Estadístico SPSS versión 15 :

Datos / Seleccionar Casos / Muestra Aleatoria - muestra

Exactamente 88 de 423 / continuar / aceptar.

De manera similar, en el Estrato 2, se enumeró el listado de la EAP de Estadística desde 1 hasta 271. Y utilizando el siguiente procedimiento en el Programa Estadístico SPSS versión 15 :

Datos / Seleccionar Casos / Muestra Aleatoria - muestra

Exactamente 57 de 271 /continuar / aceptar.

En el estrato 3, se enumeró el listado de la EAP de Investigación Operativa desde 1 hasta 457. Y utilizando el siguiente procedimiento en el Programa Estadístico SPSS versión 15 :

Datos / Seleccionar Casos / Muestra Aleatoria - muestra

Exactamente 95 de 457 / continuar / aceptar.

En el estrato 4, se enumeró el listado de la EAP de Computación Científica desde 1 hasta 291. Y utilizando el siguiente procedimiento en el Programa Estadístico SPSS versión 15:

Datos / Seleccionar Casos / Muestra Aleatoria - muestra

Exactamente 61 de 291 / continuar / aceptar.

El esquema de muestreo empleado en cada estrato fue el muestreo aleatorio simple sin reposición.

#### **4) INSTRUMENTOS DE RECOLECCIÓN DE LA INFORMACIÓN**

##### **Cuestionario**

El cuestionario utilizado se dividió en 2 partes: En una primera parte se registran los datos generales del Alumno seleccionado (Características Generales), los cuales corresponden a su edad, EAP y Ciclo que esta estudiando. La segunda parte corresponde a las características específicas

relacionado al tema de las preguntas delicadas, en donde se registra la respuesta del entrevistado. El cuestionario se muestra en el Anexo.

## **5) ACTIVIDADES DE CAMPO**

### **Personal de la encuesta y Materiales del trabajo de campo**

Para el trabajo de campo se contó con el ambiente de la Oficina 211, cuyos miembros (las profesoras de esa oficina nos concedieron ese ambiente para llevarse a cabo la encuesta) en un primer momento (parte de la entrevista), y también usamos el ambiente del Hall de la entrada principal de la FCM con el permiso de la autoridades correspondientes, para la culminación de la entrevista final. Dos profesores de la Escuela de Estadística, fueron los entrevistadores.

Para la difusión de los alumnos seleccionados, se colocó un Comunicado, con la lista de alumnos elegidos según Escuela Académico Profesional, en la puerta principal de la FCM para que ellos pudieran enterarse que habían sido elegidos, asimismo se solicitó a los Directores de cada Escuela, autorización para colocar dicho comunicado en la vitrina de cada Escuela.

La entrevista se realizó en la Oficina 211, en el horario de 10am a 8pm, la cual empezó el día 13 de octubre y finalizó el 13 de noviembre del año 2009.

Los entrevistadores recibieron charlas, sobre las definiciones utilizadas en el estudio, la explicación del cuestionario y el procedimiento aleatorizado.

Si un alumno seleccionado se presentaba; primeramente debería identificarse con su carnet universitario, con el cual el entrevistador debía identificar la Escuela y sus nombres en las listas, las cuales contenían los nombres de los alumnos seleccionados por Escuela y debía marcarlo con plumón resaltador, para que quede registrado que ese alumno ya había sido entrevistado.

El segundo paso, era explicarle verbalmente el procedimiento aleatorizado, el cual consistía de lo siguiente:

Por ejemplo para la pregunta relacionada al consumo actual de drogas, aquí tenemos diez tarjetas en las cuales hay dos tipo de preguntas, ud. seleccione una tarjeta y lea en voz baja la pregunta y diga en voz alta su respuesta, solamente su respuesta, el entrevistador anotaba la respuesta en el cuestionario en la pregunta que le correspondía (no nos interesa cual de las dos preguntas ha contestado el entrevistado solo su respuesta).

Al finalizar se le hacía entrega de un pequeño obsequio por su colaboración.

### **3.2. RESULTADOS DE LA APLICACIÓN PARA EL MODELO DE RESPUESTA ALEATORIZADA ESTRATIFICADA**

#### **3.2.1. EVALUACIÓN DEL RENDIMIENTO DE LA MUESTRA**

Se logró entrevistar a 301 alumnos seleccionados, lo cual significó una tasa de respuesta del 100%.

#### **3.2.2. CARACTERÍSTICAS DE LA POBLACIÓN INVESTIGADA**

##### **Población entrevistada**

De un total de 1442 alumnos matriculados el Semestre 2009-II, cuyas edades fluctúa entre 16 y 59 años. En el Cuadro N° III.3 se observa que, la muestra de la investigación fue de 301 alumnos, representando el 59,80% a varones y 40,20% a mujeres. También observamos que el 63,12% de alumnos seleccionados tienen edades entre 16 a 22 años; el 34,88% entre 23 a 29 años; el 0,99% entre 30 a 43 años de edad, y el 1% entre 44 a 59 años de edad.

Cuadro N° III.3  
Muestra Seleccionada por grupos de edad según sexo  
(MRA Estratificada)

GRUPOS DE EDAD	SEXO		TOTAL	%
	Femenino	Masculino		
16 - 22	76	114	<b>190</b>	63.12
23 – 29	43	62	<b>105</b>	34.88
30 – 36	0	2	<b>2</b>	0.66
37 – 43	0	1	<b>1</b>	0.33
44 – 50	0	0	<b>0</b>	0.00
51 – 59	2	1	<b>3</b>	1.00
<b>TOTAL</b>	<b>121</b>	<b>180</b>	<b>301</b>	100.00
%	40.20	59.80	100	

### 3.3. RESULTADOS DE LA APLICACIÓN PARA EL MÉTODO CONVENCIONAL

#### 3.3.1. EVALUACIÓN DEL RENDIMIENTO DE LA MUESTRA

Se logró entrevistar a 200 alumnos seleccionados, lo cual significó una tasa de respuesta del 100%.

#### 3.3.2. CARACTERÍSTICAS DE LA POBLACIÓN INVESTIGADA

##### Población entrevistada

De un total de 1442 alumnos matriculados el Semestre 2009-II, cuyas edades fluctúa entre 16 y 59 años. En el Cuadro N° III.4 se observa que, la muestra de la investigación fue de 200 alumnos, representando el 66,50% a varones y 33,50% a mujeres. También observamos que el 60% de alumnos seleccionados tienen edades entre 16 a 22 años; el 38% entre 23 a 29 años; y el 1,50% entre 30 a 43 años de edad, y el 0,50 entre 44 a 50 años de edad.



Cuadro N° III.4  
Muestra Seleccionada por grupos de edad según sexo  
(Método Convencional)

<b>GRUPOS DE</b>	<b>SEXO</b>		<b>TOTAL</b>	<b>%</b>
<b>EDAD</b>	<b>Femenino</b>	<b>Masculino</b>		
16 - 22	44	76	<b>120</b>	60.00
23 - 29	21	55	<b>76</b>	38.00
30 - 36	1	1	<b>2</b>	1.00
37 - 43	1	0	<b>1</b>	0.50
44 - 50	0	1	<b>1</b>	0.50
<b>TOTAL</b>	<b>67</b>	<b>133</b>	<b>200</b>	100.00
<b>%</b>	33.50	66.50	100	

### 3.4. COMPARACIÓN DE LOS RESULTADOS NUMÉRICOS DEL MODELOS DE RESPUESTA ALEATORIZADA ESTRATIFICADA Y EL MÉTODO CONVENCIONAL

Utilizando las formulas de los estimadores de la proporción de respondientes en la población con la pregunta sensible del MRA Estratificada de la pregunta no relacionada, presentada anteriormente, se obtuvieron los siguientes resultados:

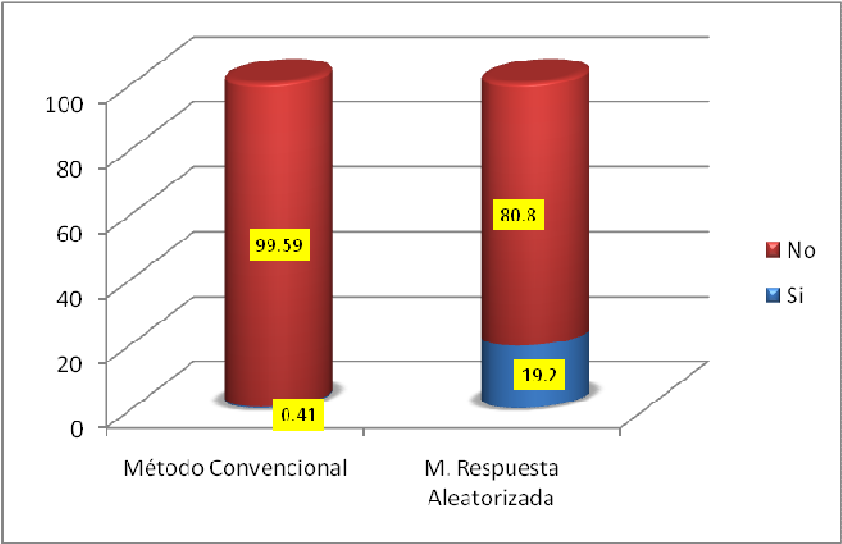
Cuadro N° III.5  
MRA Estratificada de la pregunta no relacionada y  
Método Convencional según estimación de  
Prevalencia de Drogas

<b>Método empleado</b>	<b>Prevalencia de Drogas</b>
MRA - Pregunta No Relacionada	19,20%
Método Convencional	0,41%

En el Cuadro N°III.5 observamos que utilizando el MRA Estratificada, el 19,20% de los alumnos entrevistados han consumido Pasta Básica de Cocaína (PBC) por lo menos una vez en su vida, mientras que con el Método

Convencional solo el 0,41% ha consumido PBC por lo menos una vez en su vida.

Gráfico III.1  
Efecto del tipo de método utilizado en las respuestas de Prevalencia de Drogas



Podemos observar en el Gráfico III.1, que la diferencia en los resultados, al preguntar si la persona ha consumido drogas a lo largo de su vida, es notable entre los dos métodos. Al explicarle a los entrevistados que el entrevistador no conoce la pregunta a la que responderá, aumenta disposición a decir la verdad.

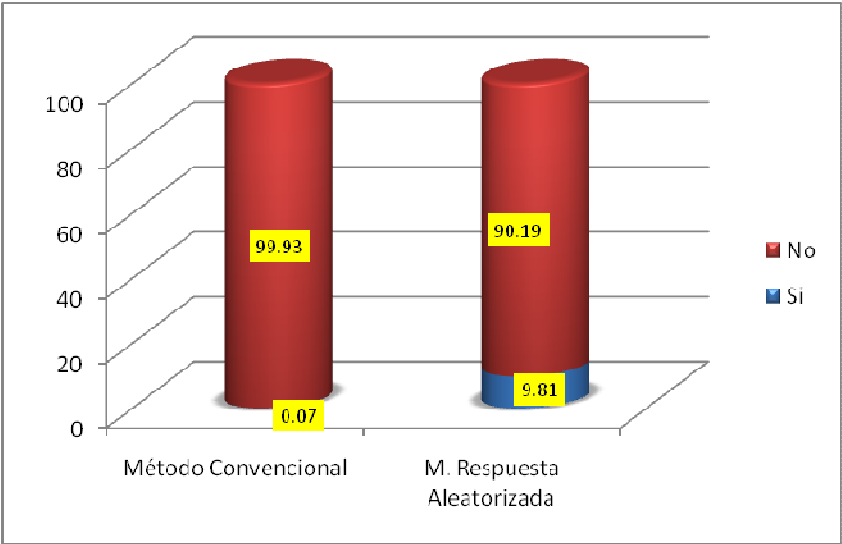
Cuadro N° III.6  
MRA Estratificada y Método Convencional según la estimación de Uso Actual de Drogas

Método empleado	Uso Actual de Drogas
MRA-Pregunta No Relacionada	9,81%
Método Convencional	0,07%

En el Cuadro N° III.6 observamos que utilizando el MRA Estratificada, el 9,81% de los alumnos entrevistados son consumidores actuales de Pasta

Básica de Cocaína (PBC), mientras que con el Método Convencional solo el 0,07% consume actualmente PBC.

Gráfico III.2  
Efecto del tipo de método utilizado en las respuestas de  
Uso actual de Drogas



En el Gráfico III.2, vemos que mediante el MRA Estratificada se obtiene un mayor porcentaje de alumnos que actualmente consume drogas, lo que no nos mostraba el Método Convencional. Gracias al MRA nos damos cuenta que debemos tomar medidas de orientación, con respecto al consumo de drogas, a los alumnos de la Facultad de Ciencias Matemáticas de la UNMSM.

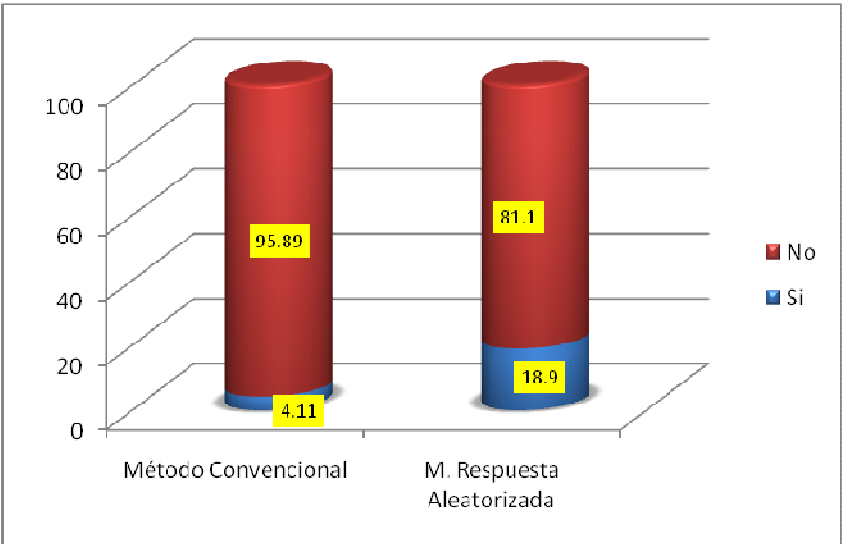
Cuadro N° III.7  
MRA Estratificada y Método Convencional según  
la estimación de Prevalencia de Sexo

Método empleado	Prevalencia de Sexo
MRA- Pregunta No Relacionada	18,9%
Método Convencional	4,11%

En el Cuadro N° III.7 observamos que utilizando el MRA estratificada, el 18,9% de los alumnos entrevistados ha mantenido relaciones sexuales con más de dos personas a lo largo de su vida, mientras que con el Método

Convencional solo el 4,11% ha mantenido relaciones sexuales con más de dos personas a lo largo de su vida.

Gráfico III.3  
Efecto del tipo de método utilizado en las respuestas de  
Prevalencia de Sexo



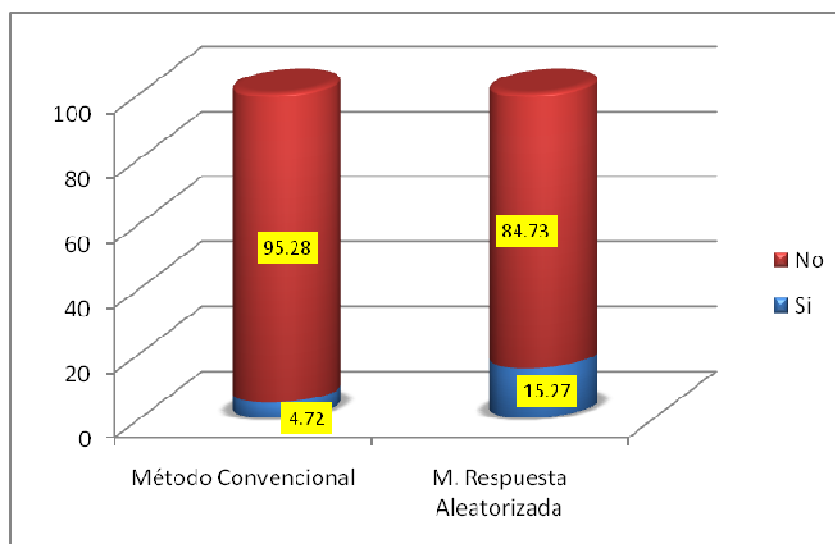
En el Gráfico III.3, vemos de igual manera que el MRA Estratificada, obtiene mejores resultados que el Método Convencional, en éste caso también se podrían tomar medidas en la Facultad de Ciencias Matemáticas, en cuanto a planificación familiar.

Cuadro N° III.8  
MRA Estratificada y Método Convencional según la estimación  
de porcentaje de alumnos que Consumen Alcohol

Método Convencional	Consumo Alcohol
MRA - Pregunta No Relacionada	15,27%
Método Convencional	4,72%

En el Cuadro N°III.8 observamos que utilizando el MRA estratificada, el 15,27% de los alumnos entrevistados consume alcohol todos los fines de semana, mientras que con el Método Convencional solo el 4,72% consume alcohol todos los fines de semana.

Gráfico III.4  
Efecto del tipo de método utilizado en las respuestas de  
Consumo de Alcohol



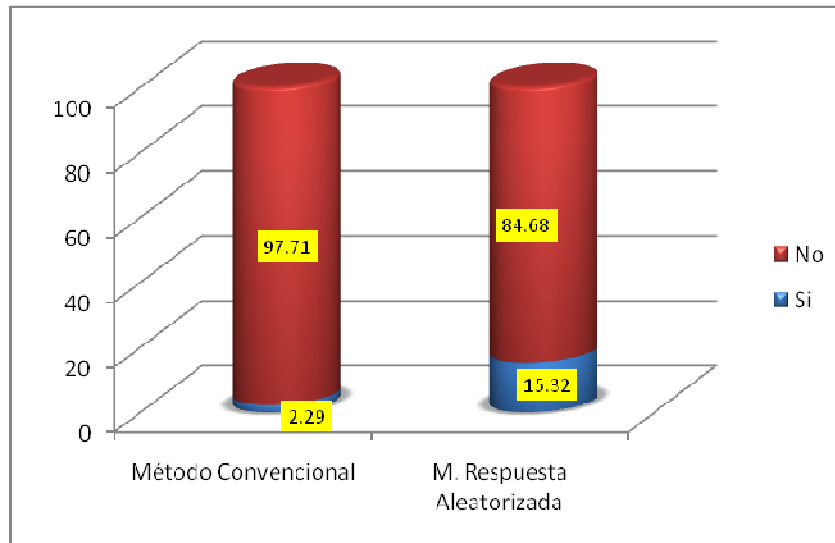
El Gráfico III.4 al igual que los demás gráficos muestra que el MRA Estratificada obtiene porcentajes mayores que el Método Convencional. Al parecer el consumo de alcohol en la Facultad de Ciencias Matemáticas es frecuente para un 15.27% de su población.

Cuadro N° III.9  
MRA Estratificada y Método Convencional según la  
estimación de Prevalencia de Hurto

Método empleado	Prevalencia Hurto
MRA - Pregunta No Relacionada	15,32%
Método convencional	2,29%

En el Cuadro N° III.9, observamos que utilizando el MRA Estratificada, el 15,32% de los alumnos entrevistados se ha llevado o ha consumido sin pagar algún producto de algún supermercado, mientras que con el Método Convencional solo el 2,29% se ha llevado o ha consumido sin pagar algún producto de algún supermercado.

Gráfico 5  
Efecto del tipo de método utilizado en las respuestas de Hurto



En el Gráfico III.5, observamos que utilizando el Método Convencional, no nos da cifras tan alarmantes como el MRA Estratificada, donde podemos tomar conciencia que pueden haber alumnos con malos hábitos y tomar medidas preventivas en salvaguarda de nuestras pertenencias y las de la Facultad.

### 3.5. PRUEBA DE HIPÓTESIS PARA LOS PARÁMETROS EN EL MODELO DE RESPUESTA ALEATORIZADA ESTRATIFICADA Y EL MÉTODO CONVENCIONAL

Ahora probaremos que los parámetros considerados en el MRA Estratificada y el Método Convencional son diferentes, en este caso cada parámetro de las cinco preguntas delicadas consideradas en la presente investigación utilizando el MRA Estratificada resulta mayor que el Método Convencional.

## Prueba de hipótesis para la “Prevalencia de Vida de Drogas”

### 1. Hipótesis

$H_0 : \pi_S = \pi_C$  : Los parámetros de “Prevalencia de Vida en Drogas” considerados son iguales

$H_1 : \pi_S > \pi_C$  : El parámetro de “Prevalencia de Vida en Drogas” considerando el MRA es mayor que el Método Convencional

### 2. Estadística de Prueba

$$Z_c = \frac{P_s - P_c}{\sqrt{P(1-P)(1/n_1 + 1/n_2)}}$$

$$P = \frac{n_s P_s + n_c P_c}{n_s + n_c}$$

Bajo la hipótesis nula tiene una distribución aproximadamente  $Z_c$  se aproxima a la distribución  $N(0,1)$  a medida que  $n$  va aumentando

### 3. Cálculo de la Estadística de Prueba

$$P_s = 0,192, \quad P_c = 0,041, \quad n_s = 301 \quad \text{y} \quad n_c = 200$$

$$P = \frac{n_s P_s + n_c P_c}{n_s + n_c} = \frac{301(0,192) + 200(0,041)}{301 + 200} = 0,13$$

### 4. Regla de decisión

$$Z_c = \frac{P_s - P_c}{\sqrt{P(1-P)(1/n_1 + 1/n_2)}} = \frac{0,192 - 0,041}{\sqrt{0,13(1-0,13)(1/301 + 1/200)}} = 4,92$$

Para un nivel de significancia del 5%,  $Z_{1-\alpha} = Z_{0,95} = 1,645$ , luego como  $Z_c = 4,92 > 1,645$ , rechazamos  $H_0$ , y concluimos que el parámetro de “Prevalencia de Vida en Drogas” considerando el MRA es mayor que el Método Convencional.

## Prueba de hipótesis para la “Uso Actual de Drogas”

### 1. Hipótesis

$H_0 : \pi_S = \pi_C$  : Los parámetros de “Consumo Actual de Drogas” considerados son iguales

$H_1 : \pi_S > \pi_C$  : El parámetro de “Consumo Actual de Drogas” considerando el MRA es mayor que el Método Convencional

### 2. Estadística de Prueba

$$Z_c = \frac{P_s - P_c}{\sqrt{P(1-P)(1/n_1 + 1/n_2)}}$$
$$P = \frac{n_s P_s + n_c P_c}{n_s + n_c}$$

Bajo la hipótesis nula tiene una distribución aproximadamente  $Z_c$  se aproxima a la distribución  $N(0,1)$  a medida que  $n$  va aumentando

### 3. Cálculo de la Estadística de Prueba

$$P_s = 0,981, P_c = 0,0007, n_s = 301 \text{ y } n_c = 200$$

$$P = \frac{n_s P_s + n_c P_c}{n_s + n_c} = \frac{301(0,981) + 200(0,0007)}{301 + 200} = 0,59$$

$$Z_c = \frac{P_s - P_c}{\sqrt{P(1-P)(1/n_1 + 1/n_2)}} = \frac{0,981 - 0,0007}{\sqrt{0,59(1-0,59)(1/301 + 1/200)}} = 21,84$$

### 4. Regla de decisión

Para un nivel de significancia del 5%,  $Z_{1-\alpha} = Z_{0,95} = 1,645$ , luego como  $Z_c = 21,84 > 1,645$ , rechazamos  $H_0$ , y concluimos que el parámetro de “Consumo Actual de Drogas” considerando el MRA es mayor que el Método Convencional.



## Prueba de hipótesis para la “Prevalencia de Sexo”

### 1. Hipótesis

$H_0 : \pi_s = \pi_c$  : Los parámetros de “Prevalencia de sexo” considerados son iguales

$H_1 : \pi_s > \pi_c$  : El parámetro de “Prevalencia de sexo” considerando el MRA es mayor que el Método Convencional

### 2. Estadística de Prueba

$$Z_c = \frac{P_s - P_c}{\sqrt{P(1-P)(1/n_1 + 1/n_2)}}$$

$$P = \frac{n_s P_s + n_c P_c}{n_s + n_c}$$

Bajo la hipótesis nula tiene una distribución aproximadamente  $Z_c$  se aproxima a la distribución  $N(0,1)$  a medida que  $n$  va aumentando

### 3. Cálculo de la Estadística de Prueba

$$P_s = 0,189, \quad P_c = 0,0411, \quad n_s = 301 \quad \text{y} \quad n_c = 200$$

$$P = \frac{n_s P_s + n_c P_c}{n_s + n_c} = \frac{301(0,189) + 200(0,0411)}{301 + 200} = 0,13$$

$$Z_c = \frac{P_s - P_c}{\sqrt{P(1-P)(1/n_1 + 1/n_2)}} = \frac{0,189 - 0,0411}{\sqrt{0,13(1-0,13)(1/301 + 1/200)}} = 4,14$$

### 4. Regla de decisión

Para un nivel de significancia del 5%,  $Z_{1-\alpha} = Z_{0,95} = 1,645$ , luego como  $Z_c = 4,14 > 1,645$ , rechazamos  $H_0$ , y concluimos que las proporciones de “Prevalencia de sexo” estimadas por el MRA es mayor que el Método Convencional.

## Prueba de hipótesis para la “Consumo Actual de Alcohol”

### 1. Hipótesis

$H_0 : \pi_S = \pi_C$  : Los parámetros de “Consumo Actual de Alcohol” considerados son iguales

$H_1 : \pi_S > \pi_C$  : El parámetro de “Consumo Actual de Alcohol” considerando el MRA es mayor que el Método Convencional

### 2. Estadística de Prueba

$$Z_c = \frac{P_s - P_c}{\sqrt{P(1-P)(1/n_1 + 1/n_2)}}$$

$$P = \frac{n_s P_s + n_c P_c}{n_s + n_c}$$

Bajo la hipótesis nula tiene una distribución aproximadamente  $Z_c$  se aproxima a la distribución  $N(0,1)$  a medida que  $n$  va aumentando

### 3. Cálculo de la Estadística de Prueba

$$P_s = 0,1527, \quad P_c = 0,0472, \quad n_s = 301 \quad \text{y} \quad n_c = 200$$

$$P = \frac{n_s P_s + n_c P_c}{n_s + n_c} = \frac{301(0,1527) + 200(0,0472)}{301 + 200} = 0,11$$

$$Z_c = \frac{P_s - P_c}{\sqrt{P(1-P)(1/n_1 + 1/n_2)}} = \frac{0,1527 - 0,0472}{\sqrt{0,11(1-0,11)(1/301 + 1/200)}} = 3,70$$

### 4. Regla de decisión

Para un nivel de significancia del 5%,  $Z_{1-\alpha} = Z_{0,95} = 1,645$ , luego como  $Z_c = 3,70 > 1,645$ , rechazamos  $H_0$ , y concluimos que las proporciones de “Consumo Actual de Alcohol” estimadas por el MRA es mayor que el Método Convencional.

## Prueba de hipótesis para la “Prevalencia de Hurto”

### 1. Hipótesis

$H_0 : \pi_s = \pi_c$ : Los parámetros de “Prevalencia de Hurto” considerados son iguales

$H_1 : \pi_s > \pi_c$ : El parámetro de “Prevalencia de Hurto” considerando el MRA es mayor que el Método Convencional

### 2. Estadística de Prueba

$$Z_c = \frac{P_s - P_c}{\sqrt{P(1-P)(1/n_1 + 1/n_2)}}$$

$$P = \frac{n_s P_s + n_c P_c}{n_s + n_c}$$

Bajo la hipótesis nula tiene una distribución aproximadamente  $Z_c$  se aproxima a la distribución  $N(0,1)$  a medida que  $n$  va aumentando

### 3. Cálculo de la Estadística de Prueba

$$P_s = 0,1532, \quad P_c = 0,0229, \quad n_s = 301 \quad \text{y} \quad n_c = 200$$

$$P = \frac{n_s P_s + n_c P_c}{n_s + n_c} = \frac{301(0,1532) + 200(0,0229)}{301 + 200} = 0,10$$

$$Z_c = \frac{P_s - P_c}{\sqrt{P(1-P)(1/n_1 + 1/n_2)}} = \frac{0,1532 - 0,0229}{\sqrt{0,11(1-0,11)(1/301 + 1/200)}} = 4,76$$

### 4. Regla de decisión

Para un nivel de significancia del 5%,  $Z_{1-\alpha} = Z_{0,95} = 1,645$ , luego como  $Z_c = 4,76 > 1,645$ , rechazamos  $H_0$ , y concluimos que el parámetro de “Consumo Actual de Alcohol” estimadas por el MRA es mayor que el Método Convencional.

## **CONCLUSIONES Y RECOMENDACIONES**

1. La aplicación de la técnica aleatorizada – el MRA Estratificada de la pregunta no relacionada – permitió poner en práctica una técnica de muestreo, que resulta eficiente en encuestas con preguntas delicadas.
2. La estimación de “Prevalencia de Droga” de la sustancia psicoactiva Pasta Básica de Cocaína con la Técnica de Respuesta Aleatorizada es de 19,20%; la estimación de “Uso Actual” de la Pasta Básica de Cocaína fue de 9,81%; la estimación de “Prevalencia de Sexo” fue de 18,9% ,la estimación de “Consumo de Alcohol” fue de 15,27% y la estimación de “Prevalencia de hurto” fue de 15,32% resultados mayores que el Método Convencional de entrevista directa.
3. De la aplicación de las Pruebas de Hipótesis, tanto para la “Prevalencia de Vida” y “Uso Actual” de la PBC, podemos concluir que las proporciones poblacionales de “Prevalencia de Vida” y “Uso Actual” de PBC por la técnica del MRA estratificada es mayor que el Método Convencional de entrevista directa, a un nivel de significancia del 5%.
4. De la aplicación de las Pruebas de hipótesis, tanto para la “Prevalencia de Sexo”, “Consumo Actual de Alcohol” y “Prevalencia de Hurto”, por la técnica del MRA Estratificada resulto mayor que el Método Convencional de entrevista directa, a un nivel de significancia del 5%.
5. Las desventajas de la Técnica de Respuesta Aleatorizada están en el hecho de que los gastos en la capacitación de los entrevistadores (el entrenamiento en la técnica) y el tiempo que requiere la entrevista para

explicar la técnica al entrevistado es más alta con respecto al Método Convencional.

6. Se recomienda seguir experimentando estos modelos, y sobre todo el Modelo de Simmons en muestras más grandes y en estudios donde el tema contenga efectivamente preguntas altamente sensibles o muy comprometedoras.

## **BIBLIOGRAFIA**

1. CAMBELL, C. and JOINER, B. "How to get the answer without being sure you've asked the question". The American Statistician, Diciembre, 1973, 229-231.
2. CHAUDHIRI, A; MUKERJEE, R. "Randomized response: theory and techniques". New York. Marcel, 1983.161p.
3. COCHRAN, G.W.Técnicas de muestreo: México: Continental, 1977. 513p.
4. DES, R. Teoría del muestreo. México: Fondo de Cultura Económica, 1980. 305p.
5. GREENBERG, B. G. et al. The unrelated question randomized response model: theorical frame work. J. Am. Stat. Assoc., Alexandria, v.64, p. 520 - 539, 1969.
6. GREENBERG, B. G. et al. Aplication on the randomized response technique in obtaining quantitative data. J. Am. Stat. Assoc., Alexandria, v. 71, n.353-, p72-73, p.980-83,1975
7. FOX, J.A: ; Tracy, P.E. Randomized Response: A Method for sensitive Surveys. Sage University Paper on Quantitative Applications in the Social Sciencies, 58. Beverly Hills: Sage Publications, 1986.
8. KIM, J-M. and ELAM, M.E. Comparison and analysis of stratified randomized response models. Joint Statistical Meetings. Section on Surveys Research Methods. J. Am. Stat. Assoc., New York, 2158-2161. 2003.
9. KIM, J-M. and ELAM, M.E.A stratified unrelated question randomized response model. Kournal of Statistical Planning and Inference, in reviw. 2003a.

10. KIM, J-M. and ELAM, M.E. A two-stage stratified Warner's randomized response model using optimal allocation. *Metrika*, in review. 2003b.
11. KIM, J-M. and ELAM, M.E. A stratified Warner's randomized response model. *Journal of Statistical Planning and Inference*, in press. 2003.
12. LANKE, S. On the choice of the unrelated question in Simmons version of randomized response model. *J. Stat. Assoc.*, New York, v.70, p.80-83, 1975.
13. MANN, C.R. Tell me your secret - random response methodology. *Advertising y marketing research professionals*, jun. 1999.
14. MOORS, J.J. A Optimization of the unrelated question randomized response model. *J. Stat. Assoc.*, New York, v. 66, n. 361, p- 627-629, 1971.
15. WARNER, S.L. Randomized response: A survey technique for elimination evasive answer bias. *J. Stat. Assoc.*, New York, v. 60, p.63-69, 1965.

# ANEXOS

UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS  
Facultad de Ciencias Matemáticas  
Escuela Académico Profesional de Estadística

## CUESTIONARIO SOBRE LA APLICACIÓN DEL MODELOS DE RESPUESTA ALEATORIZADA ESTRATIFICADA

Nº de Cuestionario

--	--	--

### I. CARACTERÍSTICAS GENERALES

1. Edad : \_\_\_\_\_

2. Sexo

1. Femenino

2. Masculino

3. E.A.P: \_\_\_\_\_

4. CICLO:

1) I 2) II 3) III 4) IV 5) V 6) VI 7) VII 8) VIII 9) IX 10) X

### II. CARACTERÍSTICAS ESPECÍFICAS

Ud. ha sido elegido al azar entre los alumnos de su Facultad para responder 5 preguntas. Quisiera explicarle el procedimiento (entrevistador presente las tarjetas). En cada grupo hay dos tipos de tarjetas con diferente contenido cada una; saque una y sin mostrarme su contenido léala en silencio y responda en voz alta, luego junte la tarjeta seleccionada con las demás de su grupo. Igualmente proceda con el otro grupo de tarjetas y así sucesivamente...

5.

DROGAS - RESULTADOS					OBSERVACIONES
PREVALENCIA	RESPONDE .....	SI	1		
(TARJETA BLANCA)	RESPONDE .....	NO	2		
USO ACTUAL	RESPONDE .....	SI	1		
(TARJETA CREMA)	RESPONDE .....	NO	2		

6.

SEXO - RESULTADOS					OBSERVACIONES
PREVALENCIA	RESPONDE .....	SI	1		
(TARJETA ROSA)	RESPONDE .....	NO	2		

7.

ALCOHOL - RESULTADOS					OBSERVACIONES
CONSUMO ACTUAL	RESPONDE .....	SI	1		
(TARJETA CELESTE)	RESPONDE .....	NO	2		

8.

HURTO - RESULTADOS					OBSERVACIONES
PREVALENCIA	RESPONDE .....	SI	1		
(TARJETA ANARANJADA)	RESPONDE .....	NO	2		

Nombre del entrevistador: \_\_\_\_\_

Fecha : \_\_\_\_\_



UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS  
Facultad de Ciencias Matemáticas  
Departamento Académico de Estadística

CUESTIONARIO SOBRE EL COMPORTAMIENTO DE LOS ALUMNOS DE LA  
UNMSM FRENTE A LAS DROGAS, ALCOHOL Y SEXO

Nº de Cuestionario

--	--	--

I. CARACTERÍSTICAS GENERALES DEL ALUMNO SELECCIONADO

1. Edad : \_\_\_\_\_

2. Sexo

1. Femenino

2. Masculino

FACULTAD : \_\_\_\_\_

3. E.A.P: \_\_\_\_\_

4. CICLO:

1) I 2) II 3) III 4) IV 5) V 6) VI 7) VII 8) VIII 9) IX 10) X

II. CARACTERÍSTICAS ESPECÍFICAS

5. ¿Durante tu vida has consumido, pasta básica de cocaína (pasta, pay o pastel) por lo menos una vez?

1. Si

2. No

6. Actualmente eres consumidor de pasta básica de cocaína (pasta, pay o pastel)?

1. Si

2. No

7. ¿Has mantenido relaciones sexuales con más de dos personas a lo largo de tu vida?

1. Si

2. No

8. ¿Consumes alcohol (cerveza, vino, sangría, etc) con frecuencia? (casi todos los fines de semana)

1. Si

2. No

Nombre del entrevistador: \_\_\_\_\_

Fecha : \_\_\_\_\_